THE HECKMAN BINDERY, INC.
North Manchester, Indiana

KRL

H or V	JUST	FONT	SLOT	TITLE
H	CC	1W	22	BEBR
			21	FACULTY
			20	WORKING
			19	PAPER
H	CC	1W	8	1989
			7	NO.1555-1572
H	CC	1W		330
				B385<"CV">
				no.1555-1572
				cop.2
				<IMPRINT>
H	CC	7W		U. of ILL.
				LIBRARY
				URBANA

BINDING COPY

PERIODICAL ☐ CUSTOM ☐ STANDARD ☐ ECONOMY ☐ THESIS | NO VOLS THIS TITLE | LEAD ATTACH

BOOK ☐ CUSTOM ☐ MUSIC ☐ ECONOMY | AUTH 1ST ☐

ACCOUNT: 66672 | LIBRARY: 001 | NEW | RUB OR SAMPLE | TITLE I.D.: 6632 | COLOR: FOIL: WHI | MATERIAL: 488

ACCOUNT NAME: UNIV OF ILLINOIS

ISSN.

ACCOUNT INTERNAL I.D.: B01912400

I.D. #2: STX3

NOTES | BINDING FREQUENCY | WHEEL: 1 | SYS. I.D.: 3 | 39254

COLLATING: 35

ADDITIONAL INSTRUCTIONS

Dept=STX3 Lot=#20 Item=143 MNM-1ZY4
1CR25T3CR MARK BY # B4 91

SEP. SHEETS	PTS. BD. PAPER	TAPE STUBS	CLOTH EXT	FILLER: GUM	STUB

LEAF ATTACH.

POCKETS: PAPER | BUCK | CLOTH

SPECIAL PREP.

INSERT MAT.

ACCOUNT LOT NO.: #20

ACCOUNT PIECE NO: 143

JOB NO: HV363

PIECE NO: 85

PRODUCT TYPE: 11

HEIGHT: 11.2

GROUP CARD: 1

VOL. THIS TITLE

COVER SIZE: X

001247549

BEBR
FACULTY WORKING
PAPER NO. 89-1568

Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis

Anil K. Bera
Sangkyu Lee

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 89-1568

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

May 1989

Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis

Anil K. Bera, Associate Professor
Department of Economics

Sangkyu Lee, Graduate Student
Department of Economics

Revised: May 1989

Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis

Anil K. Bera and Sangkyu Lee
University of Illinois at Urbana-Champaign

Abstract: We apply the White's information matrix (IM) test to the linear regression model with autocorrelated errors. A special case of one component of the test is found to be identical to the Engle's Lagrange multiplier (LM) test for autoregressive conditional heteroscedasticity ($ARCH$). Given Chesher's interpretation of the IM test as a test for parameter heterogeneity, this establishes a connection among the IM test, $ARCH$ and parameter variation. This also enables us to specify conditional heteroskedasticity in a more general and convenient way. Other interesting byproducts of our analysis are tests for the variation in conditional and unconditional skewness which we call tests for "heteroskewcity".

JEL Classification No. 210 / Key Words: Autoregressive Conditional Heteroskedasticity; Information Matrix Test; Lagrange Multiplier Test; Parameter Heterogeneity.

* Correspondence should be addressed to *Anil K. Bera*, Department of Economics, University of Illinois at Urbana–Champaign, 1206 S. Sixth Street, Champaign, IL 61801, USA.

1. INTRODUCTION

In a pioneering article, White (1982) suggested the information matrix (IM) test as a general test for model specification. In recent years, this test has received a lot of attention. In particular, Chesher (1984) demonstrated that this test can be viewed as a Lagrange multiplier (LM) test for specification error against the alternative of parameter heterogeneity. As a byproduct of this analysis, Chesher (1983) and Lancaster (1984) provided a "nR^2" version of the IM test. An application of the IM test tò the linear regression model by Hall (1987) led to a very interesting result that the test decomposed asymptotically into three components, one testing heteroskedasticity and the other two testing some forms of normality. Engle (1982), in an apparently unrelated influential paper, introduced the autoregressive conditional heteroskedasticity ($ARCH$) model which characterizes explicitly the conditional variance of the regression disturbances. He also suggested an LM test for $ARCH$. The purpose of this paper is to establish a connection among the IM test, parameter heterogeneity and $ARCH$.

An important finding by Hall (1987) was that the components of the IM test are insensitive to serial correlation. Hall also commented " had our original specification included first order autoregressive errors, then the IM test does not decompose asymptotically into the sum of our original three component test ... plus the LM test against first-order serial correlation. In this more general framework the indicator vector no longer has a block diagonal covariance matrix due to the inclusion of the autoregressive coefficient in the parameter vector." (p.262). In the next section, we start with a linear regression model with autoregressive (AR) errors and apply the IM test to it. The indicator vector is found to have a block diagonal covariance matrix. And as the null model now has more parameters, naturally we get a few extra components in the IM test. From the additional components of the statistic, we can also obtain the Engle's LM test for $ARCH$ as a special case. The implication of this result is discussed in detail in section 3. Given Chesher's interpretation of the IM test as a test for parameter heterogeneity or random coefficient, it is now easy to give a random coefficient interpretation to $ARCH$. This fact has been noted recently by

several authors [see, e.g., Tsay (1987)]. This provides us with a convenient framework to extend $ARCH$ so that interaction factor between past residuals could also be considered and as a consequence we suggest an augmented $ARCH(AARCH)$ model. In this section, we also explore the connection between $ARCH$ and overdispersion. The last section of the paper contains some concluding remarks.

2. THE IM TEST FOR THE LINEAR REGRESSION MODEL WITH AR ERRORS

We consider the linear regression model

$$y_t = x_t'\beta + \varepsilon_t \tag{1}$$

where y_t is the t-th observation on the dependent variable, x_t is a $k \times 1$ vector of *fixed* regressors and the ε_t are assumed to follow a stationary $AR(p)$ process

$$\varepsilon_t = \sum_{j=1}^{p} \phi_j \varepsilon_{t-j} + u_t \tag{2}$$

with $u_t \sim NIID(0, \sigma_u^2)$. We will write this $AR(p)$ process as $\varepsilon_t = \underline{\varepsilon}_t'\phi + u_t$ where $\underline{\varepsilon}_t = (\varepsilon_{t-1}, \cdots, \varepsilon_{t-p})'$ and $\phi = (\phi_1, \cdots, \phi_p)'$. Assuming that $\underline{\varepsilon}_t$ is given, the log- likelihood function for this model can be written as

$$L(\theta) = \sum_{t=1}^{n} l_t(\theta) = -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_{t=1}^{n}(\varepsilon_t - \underline{\varepsilon}_t'\phi)^2$$

where $\theta = (\beta', \phi', \sigma_u^2)'$ is a $(k+p+1) \times 1$ vector of parameters. Note that $(\varepsilon_t - \underline{\varepsilon}_t'\phi)$ involves β since $\varepsilon_t - \underline{\varepsilon}_t'\phi = (y_t - \underline{y}_t'\phi) - (x_t - \underline{x}_t'\phi)'\beta$, where $\underline{y}_t = (y_{t-1}, \cdots, y_{t-p})'$ and $\underline{x}_t = (x_{t-1}, \cdots, x_{t-p})'$.

Let $\hat{\theta}$ denote the maximum likelihood estimate (MLE) of θ. Then White's IM test is constructed based on

$$d(\hat{\theta}) = vech \quad C(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} d_t(\hat{\theta}) \qquad (say)$$

where

$$C(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\hat{\theta})}{\partial\theta\partial\theta'} + \left(\frac{\partial l_t(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l_t(\hat{\theta})}{\partial\theta}\right)'\right] = A(\hat{\theta}) + B(\hat{\theta}) \qquad (say)$$

A consistent estimator of the variance matrix of $d(\hat{\theta})$ is [see White (1982, p.11)]

$$V(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} a_t(\hat{\theta})a_t'(\hat{\theta}) \tag{3}$$

where $a_t(\hat{\theta}) = d_t(\hat{\theta}) - \nabla d(\hat{\theta})A(\hat{\theta})^{-1}\nabla l_t(\hat{\theta})$ with $\nabla d(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial d_t(\hat{\theta})}{\partial\theta}$ and $\nabla l_t(\hat{\theta}) = \frac{\partial l_t(\hat{\theta})}{\partial\theta}$. Then White's IM test takes the form of

$$T_W = nd'(\hat{\theta})V_a(\hat{\theta})^{-1}d(\hat{\theta}) \tag{4}$$

When the model (1) is correct, T_W follows an asymptotic χ^2 with $\frac{q(q+1)}{2}$ degrees of freedom. Here it should be noted that White (1982) derived the IM test for IID observations. However, as shown in White (1987), the IM equality holds under fairly general conditions. For our autoregressive case, mixing conditions stated in White (1987) are satisfied, and therefore the IM test remains valid.

After some algebra and rearranging the terms in $d(\hat{\theta})$, we can write (for algebraic derivations, see the Appendices A and B), suppressing θ such that $\hat{d}$ for $d(\hat{\theta})$,

$$\hat{d} = (\hat{d}_1', \hat{d}_2', \hat{d}_3', \hat{d}_4', \hat{d}_5', \hat{d}_6')' \tag{5}$$

where $\hat{d}_1$ is a $\frac{k(k+1)}{2} \times 1$ vector of the difference of two estimates of the variance of $\hat{\beta}$, $\hat{d}_2$ is a $\frac{p(p+1)}{2} \times 1$ vector of the difference of two estimates of the variance of $\hat{\phi}$, $\hat{d}_3$ is a scalar of the difference of two estimates of the variance of $\hat{\sigma}_u^2$, $\hat{d}_4$ is a $kp \times 1$ vector of the difference of two estimates of the covariance between $\hat{\beta}$ and $\hat{\phi}$, $\hat{d}_5$ is a $k \times 1$ vector of the difference of two estimates of the covariance between $\hat{\beta}$ and $\hat{\sigma}_u^2$, $\hat{d}_6$ is a $p \times 1$ vector of the difference of two estimates of the covariance between $\hat{\phi}$ and $\hat{\sigma}_u^2$, and the typical elements of d_i, $\quad i = 1, 2, \cdots, 6$, are given below

$$\hat{d}_1 : \qquad \left[\frac{1}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}(\hat{u}_t^2 - \hat{\sigma}_u^2)(x_{ti} - \underline{x}_{ti}'\hat{\phi})(x_{tj} - \underline{x}_{tj}'\hat{\phi})\right]_{i,j=1,2,\cdots,k; \quad i\leq j}$$

$$\hat{d}_2 : \quad \left[\frac{1}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}(\hat{u}_t^2 - \hat{\sigma}_u^2)\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}\right]_{i,j=1,2,\cdots,p; \quad i\leq j}$$

$$\hat{d}_3 : \quad \left[\frac{1}{4n\hat{\sigma}_u^8}\sum_{t=1}^{n}(\hat{u}_t^4 - 3\hat{\sigma}_u^4)\right]$$

$$\hat{d}_4 : \quad \left[\frac{1}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}(\hat{u}_t^2 - \hat{\sigma}_u^2)(x_{ti} - \underline{x}'_{ti}\hat{\phi})\hat{\varepsilon}_{t-j}\right]_{i=1,2,\cdots,k; \quad j=1,2,\cdots,p}$$

$$\hat{d}_5 : \quad \left[\frac{1}{2n\hat{\sigma}_u^6}\sum_{t=1}^{n}\hat{u}_t^3(x_{ti} - \underline{x}'_{ti}\hat{\phi})\right]_{i=1,2,\cdots,k}$$

$$\hat{d}_6 : \quad \left[\frac{1}{2n\hat{\sigma}_u^6}\sum_{t=1}^{n}\hat{u}_t^3\hat{\varepsilon}_{t-i}\right]_{i=1,2,\cdots,p}$$

Our expressions for $\hat{d}_1, \hat{d}_3$ and $\hat{d}_5$ are identical to those of $\triangle_1, \triangle_3$ and $\triangle_2$ of Hall (1987, pp. 259-260) if we put $\hat{\phi} = 0$. If it is desirable to test only in certain direction, we can premultiply $\hat{d}$ by a selection matrix whose elements are either zero or unity [see White (1982, pp.9-10) and Hall (1987, p.258)].

Now to obtain the IM test statistic, all we need is to derive the variance matrix of $\hat{d}$. We find that the variance matrix is block diagonal (for detailed derivation, see the Appendix B). Denote the estimator of the variance of $\sqrt{n}\hat{d}_i$ as $V(\hat{d}_i) = \hat{V}_i, i = 1, 2, \cdots, 6$. To express the $\hat{V}_i$'s succinctly, we define the vectors whose typical elemens are described as

$$\underline{\chi}_t : \quad \left[(x_{ti} - \underline{x}'_{ti}\hat{\phi})(x_{tj} - \underline{x}'_{tj}\hat{\phi}) - \frac{1}{n}\sum_{t=1}^{n}(x_{ti} - \underline{x}'_{ti}\hat{\phi})(x_{tj} - \underline{x}'_{tj}\hat{\phi})\right]_{i,j=1,2,\cdots,k; \quad i\leq j}$$

$$\underline{\xi}_t : \quad \left[\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j} - \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}\right]_{i,j=1,2,\cdots,p; \quad i\leq j}$$

$$\underline{s}_t : \quad \left[(x_{ti} - \underline{x}'_{ti}\hat{\phi})\hat{\varepsilon}_{t-j}\right]_{i=1,2,\cdots,k; \quad j=1,2,\cdots,p}$$

$$\underline{r}_t : \quad \left[x_{ti} - \underline{x}'_{ti}\hat{\phi}\right]_{i=1,2,\cdots,k}$$

Then we have very concise forms of the $\hat{V}_i$s as follows

$$\hat{V}_1 = \frac{2}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}\underline{\chi}_t\underline{\chi}_t', \quad \hat{V}_2 = \frac{2}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}\underline{\xi}_t\underline{\xi}_t', \quad \hat{V}_3 = \frac{3}{2\hat{\sigma}_u^8}$$

$$\hat{V}_4 = \frac{2}{n\hat{\sigma}_u^4}\sum_{t=1}^{n}\underline{s}_t\underline{s}_t', \quad \hat{V}_5 = \frac{1}{2n\hat{\sigma}_u^6}\sum_{t=1}^{n}\underline{r}_t\underline{r}_t', \quad \hat{V}_6 = \frac{1}{2n\hat{\sigma}_u^6}\sum_{t=1}^{n}\underline{\hat{\varepsilon}_t\hat{\varepsilon}_t'}$$

Given the block diagonality of the variance matrix of $\hat{d}$, we can write the IM test as

$$T_W = \sum_{i=1}^{6} T_i = n\sum_{i=1}^{6}\hat{d}_i'\hat{V}_i^{-1}\hat{d}_i \tag{6}$$

that is, the derived IM test statistic is found to be decomposed as the sum of six quadratic forms. In the next section, we analyze these components of T_W in detail.

3. INTERPRETATION OF THE COMPONENTS OF THE IM TEST

Using Chesher's analysis, we can say the statistic T_1 is a test for randomness of the regression parameters in the presence of autocorrelation. If we put $\hat{\phi} = 0$, then this reduces to the White (1980)'s test for heteroskedasticity [and T_{1n} in Hall (1987, p.261)]. Recently, there have been some robustness studies of various tests for heteroskedasticity in the presence of autocorrelation [see, e.g., Epps and Epps (1977), Bera and Jarque (1982), Godfrey and Wickens (1982), Bumb and Kelejian (1983), Bera and McKenzie (1986)] and their general conclusion is that various tests for heteroskedasticity is sensitive to the presence of autocorrelation. A byproduct of our analysis is that we have a simple test for heteroskedasticity in the presence of autocorrelation. All we need to do is to modify the White test slightly. Instead of regressing the squares of the least squares residuals on the squares and cross products of x_t's, we should regress $\hat{u}_t^2$ on the squares and cross products of $(x_t - \underline{x}_t'\hat{\phi})$ after estimating the model with appropriate AR process. For example, if there is $AR(1)$ error, then the regressors should be the squares and cross products of $(x_t - \hat{\phi}_1 x_{t-1})$. Similarly, the modification of T_{2n} in Hall (1987), which is our T_5, requires

that we should replace x_t by $(x_t - \underline{x}_t'\hat{\phi})$. Our T_3 is a (kurtosis) test for normality. Here all we need is to use the conditional mean corrected residuals rather than the OLS residuals.

Let us now concentrate on the new test statistics we obtain by including ϕ in our model. The statistic T_2 tests the randomness of $\phi = (\phi_1, \phi_2, \cdots, \phi_p)'$. Suppose that the parameters of autoregressive errors are varying around a mean value with finite variances. This can be formulated as $\phi_t \sim (\phi, \Omega)$, where $\phi_t = (\phi_{1t}, \phi_{2t}, \cdots, \phi_{pt})'$. Then T_2 is the LM statistic for testing $H_0 : \Omega = 0$. Let us first consider a very special case in which $\phi = 0$ and Ω is diagonal. Under H_0 : $\hat{\phi}_1 = \hat{\phi}_2 = \cdots = \hat{\phi}_p = 0$, $\hat{u}_{t-i} = \hat{\varepsilon}_{t-i}$ $(i = 1, 2, \cdots, p)$, where the $\hat{\varepsilon}_t$ are the OLS residuals. Consequently, T_2 reduces to

$$T_2 = \frac{1}{2}\left[\sum_{t=1}^{n} \underline{\hat{u}_t^2}\left(\frac{\hat{u}_t^2}{\hat{\sigma}_u^2} - 1\right)\right]' \left[\sum_{t=1}^{n} \underline{\xi}_t \underline{\xi}_t'\right]^{-1} \left[\sum_{t=1}^{n} \underline{\hat{u}_t^2}\left(\frac{\hat{u}_t^2}{\hat{\sigma}_u^2} - 1\right)\right] \tag{7}$$

where$\underline{\hat{u}_t^2} = (\hat{u}_{t-1}^2, \hat{u}_{t-2}^2, \cdots, \hat{u}_{t-p}^2)'$and a typical element of $\underline{\xi}_t$ is now $\left(\hat{u}_{t-i}^2 - \frac{1}{n}\sum_{t=1}^{n} \hat{u}_{t-i}^2\right)$, for $i = 1, 2, \cdots, p$. This is identical to the Engle's (1982) LM statistic for testing the pth-order linear $ARCH$ disturbances, i.e., testing $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ in the $ARCH$ process specified as $Var(u_t \mid \underline{u}_t) = \sigma_u^2 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2$, where $\underline{u}_t = (u_{t-1}, u_{t-2}, \cdots, u_{t-p})'$. An asymptotically equivalent form of this statistic is nR^2 where R^2 is the coefficient of multiple determination from the regression of $\hat{u}_t^2$ on a unity and $(\hat{u}_{t-1}^2, \hat{u}_{t-2}^2, \cdots, \hat{u}_{t-p}^2)$.

From our representation of test for $ARCH$ as a test for randomness of ϕ parameters and its equivalence to one component of the IM test, the consequence of the presence of $ARCH$ is that the "usual" estimators for variance of $\hat{\phi}$ will be inconsistent if $ARCH$ is ignored. This is similar to the case that the standard variance estimator for $\hat{\beta}$ is inconsistent in the presence of unconditional heteroskedasticity. Therefore, the standard tests for autocorrelation are not valid in the presence of $ARCH$ [see, e.g., Diebold (1986)].

Now we relax the assumption of the diagonality of Ω. The structure of the test statistic will remain the same but that R^2 will be obtained by regressing $\hat{u}_t^2$ on a constant and the squares and cross products of the lagged residuals. T_2 will then be a LM statistic for

testing $H_0 : \alpha_{ij} = 0 (i \geq j = 1, 2, \cdots, p)$ in

$$Var(u_t \mid \underline{u}_t) = \sigma_u^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{ij} u_{t-i} u_{t-j} \qquad i \geq j \tag{8}$$

The above specification of conditional variance generalizes Engle's $ARCH$ model. This will be called the augmented $ARCH$ ($AARCH$) process. Properties and testing of this model is discussed in Bera and Lee (1988). Lastly, if we additionally relax the assumption of $\phi = 0$, $\hat{u}_t$ will no longer be equal to $\hat{\varepsilon}_t$ and T_2 will have to be calculated from the regression of $\hat{u}_t^2$ on a constant and the squares and cross products of $\hat{\varepsilon}_{t-i}$ $(i = 1, 2, \cdots, p)$. This will give us the LM statistics for testing $ARCH$ or $AARCH$ in the presence of autocorrelation.

From the above discussion, it is clear that Engle's $ARCH$ model can be viewed as a special case of random coefficient autoregressive model ($RCAR$). To see this more clearly, let us write equation (2) as $\varepsilon_t = \sum_{j=1}^{p} \phi_{jt}\varepsilon_{t-j} + u_t$. If it is assumed that $\phi_{jt} \sim (0, \alpha_j)$ and $cov(\phi_{jt}, \phi_{j't}) = 0$, for $j \neq j'$, then the conditional variance is given by $Var(\varepsilon_t \mid \underline{\varepsilon}_t) = \sigma_u^2 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2$. Here we observe that $ARCH$ and the above $RCAR$ models have the same first two conditional moments as mentioned in Tsay (1987) where it is called as second-order equivalence. If we further assume normality of ϕ_{jt}, then all the moments of $ARCH$ and $RCAR$ processes will be the same, e.g., for $p = 1$, the first four moments are $\mu_1 = 0$, $\mu_2 = \frac{\sigma_u^2}{1-\alpha_1}$, $\mu_3 = 0$ and $\mu_4 = \frac{3\sigma_u^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$ [see Engle (1982, p.992)]. Here we should note that calculation of moments are much easier under the $RCAR$ scheme.

Presence of $ARCH$ can also be viewed as "overdispersion" in the following sense. For simplicity, consider $\varepsilon_t = \phi_{1t}\varepsilon_{t-1} + u_t$ where the u_t's are $IID(0, 1)$. When ϕ_{1t} is fixed at μ, $Var(\varepsilon_t) = (1 - \mu^2)^{-1}$. If $\phi_{1t} \sim (\mu, \alpha_1)$ with $\frac{\alpha_1}{(1-\mu^2)} < 1$ and is independent of ε_t's and u_t's, then $Var(\varepsilon_t) = (1 - \mu^2 - \alpha_1)^{-1} \geq (1 - \mu^2)^{-1}$. It should be noted that $\frac{\alpha_1}{(1-\mu^2)} < 1$ is the stationarity condition for $ARCH$ in the presence of AR as discussed in Bera and Lee (1988). Cox (1983) suggested a test for $\alpha_1 = 0$, with overdispersion on the borderline of detectability, i.e., with local alternatives like $\alpha_1 = \frac{\tau}{\sqrt{n}}$. Under the certain regularity conditions, Cox expressed the density of u_t in the overdispersed model as

$$E\left[f_t(u; \phi)\right] = f_t(u; \mu)\left[1 + \frac{\tau}{2\sqrt{n}} h_t(u; \mu) + O(\frac{1}{n})\right] \tag{9}$$

where $h_t(u;\mu) = \left[\frac{\partial \log f_t(u;\mu)}{\partial \mu}\right]^2 + \left[\frac{\partial^2 \log f_t(u;\mu)}{\partial \mu^2}\right]$ which is the same as d_t defined earlier. As a test for overdispersion, Cox suggested to use $\sum_t h_t(u_t;\hat{\mu})$, where $\hat{\mu}$ is the MLE of μ under $\alpha_1 = 0$. This is essentially the IM test [see also Chesher (1983, fn 4)].

By comparing T_1 and T_2, we note that they test for unconditional and conditional heteroskedasticity, respectively. Given the block diagonality of covariance matrix of the IM test in our case, we can test for unconditional and conditional heteroskedasticity simultaneously simply by adding up these two statistics. The statistic T_4 is also related to T_1 and T_2. For obtaining T_4, we run the regression of $\hat{u}_t^2$ on a constant and cross products of lagged residuals and exogenous variables. This indicates that the form of heteroskedasticity under the alternative hypothesis would be

$$Var(u_t \mid \underline{\varepsilon}_t) = \sigma_u^2 + \sum_{i=1}^{n}\sum_{j=1}^{p} \delta_{ij} x_{ti} \varepsilon_{t-j} \tag{10}$$

and we test $H_0 : \delta_{ij} = 0$. This can be viewed as a form of conditional heteroskedasticity caused by the interaction between the disturbance term and the regressors. As a natural consequence, a general test statistic for heteroskedasticity would be $T_1+T_2+T_4$ which under the null hypothesis will have an asymptotic χ^2 distribution with $(k+p)(k+p+1)/2$ degrees of freedom. To get reasonable power, we will have to make a judicious selection of the regressors from the set of squares and cross products of $\{x_{ti}, i = 1, 2, \cdots, k$ and $\hat{\varepsilon}_{t-j}, j = 1, 2, \cdots, p\}$ or make some adjustment to the test statistic [see Bera (1986)].

The last two statistics T_5 and T_6 can be viewed as the statistics for testing variation in the third moment of u_t. In T_5, the variation is assumed to depend on the exogenous variables x_t and in T_6, on the lagged innovation process. In some sense, we could say that T_5 and T_6 test for unconditional and conditional "heteroskewcity", respectively. As noted in Hall (1987), the test for normality (skewness part) proposed by Bowman and Shenton (1975) and Jarque and Bera (1987) is a special case of T_5 while T_3 which tests for the variation of σ_u^2 is a pure test for kurtosis. In this connection, we conjecture that if the IM test is applied to an $ARCH$ model, that will lead to a test for "heterokurtosicity".

Here we should note that all the components of T are related to tests for the second, third and fourth moments of u_t. Therefore, using the standard IM test, we cannot test for the specification of the first moment of u_t, e.g., $E(\varepsilon_t \mid \underline{\varepsilon}_t) = 0$. By construction, the IM test is based on the difference of two estimated covariance matrices and implicitly, it tests for some parametric variation. In the context of specification tests for latent models, Gourieroux, et al (1987, p.28) noted that the IM test implicitly checks whether the model is "second-order" well specified according to the analysis of variance equation. Obviously, it is not possible to express a hypothesis regarding the mean part of the model such as $H_0 : \phi_1 = \phi_2 = \cdots = \phi_p = 0$ in equation (2) in terms of parameter variation. That is why the IM test fails to detect autocorrelation.

Lastly, we should mention that using his dynamic information matrix (DIM) test White (1987) obtained a test for $ARCH$ and Durbin-Watson type test for autocorrelation. The DIM test is based on the idea that under correct specification the derivatives of the conditional log-likelihood of the observation at each time period are martingale difference sequences. This test is not based explicitly on the difference of two variance-covariance matrix estimators. Also, it is not clear whether the DIM test could be given a test for parameter variation interpretation. Also it is not based explicitly on the difference of two variance-covariance matrix estimators.

4. CONCLUSION

Our application of the White's IM test to the linear regression model with autoregressive errors provides many interesting results. The most important result is that a special case of one component of this test is identical to the Engle's LM test for $ARCH$. Chesher's interpretation of the IM test as the test for parameter heterogeneity leads us naturally to specify the $ARCH$ processes as a random coefficient autoregressive ($RCAR$) model. From both theoretical and practical points of view, this representation of $ARCH$ is convenient and useful. As discussed in Bera and Lee (1988), we can now easily verify the stationarity condition for $ARCH$ as a special case of $RCAR$ model, study the robustness

of test for AR process in the presence of $ARCH$ and vice versa and generalize the $ARCH$ process to take account of interaction between the disturbance terms.

The difference between the unconditional and conditional heteroskedasticity is now clear. The former is related to the variation of the regression coefficients while the latter to the variation of the autoregressive parameters. A mixture of them is possible when the heteroskedasticity is caused by the interaction between exogenous variables and disturbances. We can now talk about the conditional and unconditional variation in skewness. As discussed in the last section, the standard IM test cannot test for the first moment part of the model, such as the presence of autocorrelation. However, it is worth noting that the power of the IM test lies in testing for *higher order* conditional and unconditional moments. Since economic theory, in most cases, provides information concerning the first moment only, the problem of testing for higher order moments is an important issue in econometric modeling. In that context, the IM test is a very useful tool for model specification.

APPENDIX

A: The Derivatives of the Log-likelihood Function. For our model, the vector of parameters is $\theta = (\beta', \gamma', \sigma_u^2)'$ and the log-likelihood function for the t-th observation conditional on the information set Ψ_{t-1}, in which $\underline{\varepsilon}_t = (\varepsilon_{t-1}, \cdots, \varepsilon_{t-p})'$ is contained, is given by $l_t(\theta) = -\frac{1}{2}log2\pi - \frac{1}{2}log\sigma_u^2 - \frac{1}{2\sigma_u^2}(\varepsilon_t - \underline{\varepsilon}_t'\phi)^2$. Note that $u_t = \varepsilon_t - \underline{\varepsilon}_t'\phi = (y_t - \underline{y}_t'\phi) - (x_t - \underline{x}_t'\phi)'\beta$ where $\underline{y}_t = (y_{t-1}, \cdots, y_{t-p})'$ and $\underline{x}_t = (x_{t-1}, \cdots, x_{t-p})'$. Then the first partial derivatives of $l_t(\theta)$ with respect to θ are easily obtained.The first derivatives are $\frac{\partial l_t(\theta)}{\partial \beta} = \frac{1}{\sigma_u^2}u_t(x_t - \underline{x}_t'\phi)$, $\frac{\partial l_t(\theta)}{\partial \phi} = \frac{1}{\sigma_u^2}u_t\underline{\varepsilon}_t$ and $\frac{\partial l_t(\theta)}{\partial \sigma_u^2} = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}u_t^2$. And the second derivatives are $\frac{\partial^2 l_t(\theta)}{\partial \beta \partial \beta'} = -\frac{1}{\sigma_u^2}(x_t - \underline{x}_t'\phi)(x_t - \underline{x}_t'\phi)'$, $\frac{\partial^2 l_t(\theta)}{\partial \phi \partial \phi'} = -\frac{1}{\sigma_u^2}\underline{\varepsilon}_t\underline{\varepsilon}_t'$, $\frac{\partial^2 l_t(\theta)}{\partial (\sigma_u^2)^2} = \frac{1}{2\sigma_u^4} - \frac{1}{\sigma_u^6}u_t^2$, $\frac{\partial^2 l_t(\theta)}{\partial \beta \partial \phi'} = -\frac{1}{\sigma_u^2}(x_t - \underline{x}_t'\phi)\underline{\varepsilon}_t'$, $\frac{\partial^2 l_t(\theta)}{\partial \beta \partial \sigma_u^2} = -\frac{1}{\sigma_u^4}u_t(x_t - \underline{x}_t'\phi)'$ and $\frac{\partial^2 l_t(\theta)}{\partial \phi \partial \sigma_u^2} = -\frac{1}{\sigma_u^4}u_t\underline{\varepsilon}_t'$.

B: A Consistent Covariance Matrix Estimator for the Information Matrix Test. A consistent covariance estimator for the IM test proposed by White (1982) is stated as

$$V(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} a_t(\hat{\theta})a_t(\hat{\theta})' \tag{B.1}$$

where $a_t(\hat{\theta}) = d_t(\hat{\theta}) - \nabla d(\hat{\theta})A(\hat{\theta})^{-1}\nabla l_t(\hat{\theta})$. Each component of $a_t(\hat{\theta})$ will be defined and derived one by one. Let us begin with the indicator vector $d(\hat{\theta})$ which is defined as

$$d(\hat{\theta}) = vech[C(\hat{\theta})] = vech[A(\hat{\theta}) + B(\hat{\theta})]$$

where

$$\begin{aligned} A(\hat{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]_{\theta=\hat{\theta}} \\ &= \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} -\frac{1}{\sigma_u^2}(x_t - \underline{x}_t'\phi)(x_t - \underline{x}_t'\phi)' & -\frac{1}{\sigma_u^2}(x_t - \underline{x}_t'\phi)\underline{\varepsilon}_t' & -\frac{1}{\sigma_u^4}(x_t - \underline{x}_t'\phi)u_t \\ -\frac{1}{\sigma_u^2}\underline{\varepsilon}_t(x_t - \underline{x}_t'\phi)' & -\frac{1}{\sigma_u^2}\underline{\varepsilon}_t\underline{\varepsilon}_t' & -\frac{1}{\sigma_u^4}\underline{\varepsilon}_t u_t \\ -\frac{1}{\sigma_u^4}u_t(x_t - \underline{x}_t'\phi)' & -\frac{1}{\sigma_u^4}u_t\underline{\varepsilon}_t' & \frac{1}{2\sigma_u^4} - \frac{1}{\sigma_u^6}u_t^2 \end{bmatrix}_{\theta=\hat{\theta}} \end{aligned}$$

and

$$B(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\left[\left(\frac{\partial l_t(\theta)}{\partial\theta}\right)\left(\frac{\partial l_t(\theta)}{\partial\theta}\right)'\right]_{\theta=\hat{\theta}}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \frac{1}{\sigma_u^4}u_t^2(x_t-\underline{x}_t'\phi)(x_t-\underline{x}_t'\phi)' & \frac{1}{\sigma_u^4}u_t^2(x_t-\underline{x}_t'\phi)\underline{\varepsilon}_t' & \frac{1}{2\sigma_u^6}(x_t-\underline{x}_t'\phi)(u_t^3-\sigma_u^2u_t) \\ \frac{1}{\sigma_u^4}u_t^2\underline{\varepsilon}_t(x_t-\underline{x}_t'\phi)' & \frac{1}{\sigma_u^4}u_t^2\underline{\varepsilon}_t\underline{\varepsilon}_t' & \frac{1}{2\sigma_u^6}\underline{\varepsilon}_t(u_t^3-\sigma_u^2u_t) \\ \frac{1}{2\sigma_u^6}(u_t^3-\sigma_u^2u_t)(x_t-\underline{x}_t'\phi)' & \frac{1}{2\sigma_u^6}(u_t^3-\sigma_u^2u_t)\underline{\varepsilon}_t' & \frac{1}{4\sigma_u^4}-\frac{1}{2\sigma_u^6}u_t^2+\frac{1}{4\sigma_u^8}u_t^4 \end{bmatrix}_{\theta=\hat{\theta}}$$

From $A(\hat{\theta})$ and $B(\hat{\theta})$, $C(\hat{\theta})$ is easily derived as

$$C(\hat{\theta}) = A(\hat{\theta}) + B(\hat{\theta})$$

$$= \frac{1}{n}\sum_{t=1}^{n}$$

$$\begin{bmatrix} \frac{1}{\sigma_u^4}(u_t^2-\sigma_u^2)(x_t-\underline{x}_t'\phi)(x_t-\underline{x}_t'\phi)' & \frac{1}{\sigma_u^4}(u_t^2-\sigma_u^2)(x_t-\underline{x}_t'\phi)\underline{\varepsilon}_t' & \frac{1}{2\sigma_u^6}(x_t-\underline{x}_t'\phi)(u_t^3-3\sigma_u^2u_t) \\ \frac{1}{\sigma_u^4}(u_t^2-\sigma_u^2)\underline{\varepsilon}_t(x_t-\underline{x}_t'\phi)' & \frac{1}{\sigma_u^4}(u_t^2-\sigma_u^2)\underline{\varepsilon}_t\underline{\varepsilon}_t' & \frac{1}{2\sigma_u^6}\underline{\varepsilon}_t(u_t^3-3\sigma_u^2u_t) \\ \frac{1}{2\sigma_u^6}(u_t^3-3\sigma_u^2u_t)(x_t-\underline{x}_t'\phi)' & \frac{1}{2\sigma_u^6}(u_t^3-3\sigma_u^2u_t)\underline{\varepsilon}_t' & \frac{1}{4\sigma_u^8}(u_t^4-6\sigma_u^2u_t^2+3\sigma_u^4) \end{bmatrix}_{\theta=\hat{\theta}}$$

Now it is straightforward to obtain $d(\hat{\theta})$. For analytical convenience, we rearrange $d(\hat{\theta})$ as described in the paper. Then the first component of $a_t(\hat{\theta})$ defined from $d(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} d_t(\hat{\theta})$ can be written as

$$d_t(\hat{\theta}) = (\hat{d}_{t1}', \hat{d}_{t2}', \hat{d}_{t3}', \hat{d}_{t4}', \hat{d}_{t5}', \hat{d}_{t6}')' \tag{B.2}$$

where $\hat{d}_{t1} = [\hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_{t1}-\underline{x}_{t1}'\hat{\phi})^2, \cdots, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_{tk}-\underline{x}_{tk}'\hat{\phi})^2, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_{t1}-\underline{x}_{t1}'\hat{\phi})(x_{t2}-\underline{x}_{t2}'\hat{\phi}), \cdots, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_{t(k-1)}-\underline{x}_{t(k-1)}'\hat{\phi})(x_{tk}-\underline{x}_{tk}'\hat{\phi})]'$ is a $\frac{k(k+1)}{2}\times 1$ vector of the difference of two estimates of the variance of $\hat{\beta}$, $\hat{d}_{t2} = [\hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)\hat{\varepsilon}_{t-1}^2, \cdots, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)\hat{\varepsilon}_{t-p}^2, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-2}, \cdots, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)\hat{\varepsilon}_{t-p+1}\hat{\varepsilon}_{t-p}]'$ is a $\frac{p(p+1)}{2}\times 1$ vector of the differences of two estimates of the variance of $\hat{\phi}$, $\hat{d}_{t3} = (4\hat{\sigma}_u^8)^{-1}(\hat{u}_t^4-6\hat{\sigma}_u^2\hat{u}_t^2+3\hat{\sigma}_u^4)$ is a scalar of the difference of two estimates of the variance of $\hat{\sigma}_u^2$, $\hat{d}_{t4} = [\hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_t-\underline{x}_t'\phi)'\hat{\varepsilon}_{t-1}, \cdots, \hat{\sigma}_u^{-4}(\hat{u}_t^2-\hat{\sigma}_u^2)(x_t-\underline{x}_t'\phi)'\hat{\varepsilon}_{t-p}]'$ is a $kp\times 1$ vector of the difference of two estimates of the covariance between $\hat{\beta}$ and $\hat{\phi}$, $\hat{d}_{t5} = (2\hat{\sigma}_u^6)^{-1}(\hat{u}_t^3-3\hat{\sigma}_u^2\hat{u}_t)(x_t-\underline{x}_t'\hat{\phi})$ is a

$k \times 1$ vector of the difference of two estimates of the covariance between $\hat{\beta}$ and $\hat{\sigma}_u^2$ and finally, $\hat{d}_{t6} = [(2\hat{\sigma}_u^6)^{-1}(\hat{u}_t^3 - 3\hat{\sigma}_u^2\hat{u}_t)\hat{\varepsilon}_{t-1}, \cdots, (2\hat{\sigma}_u^6)^{-1}(\hat{u}_t^3 - 3\hat{\sigma}_u^2\hat{u}_t)\hat{\varepsilon}_{t-p}]'$ is a $p \times 1$ vector of the difference of two estimates of the covariance between $\hat{\phi}$ and $\hat{\sigma}_u^2$.

To simplify a consistent estimator, say $g(\hat{\theta})$ which takes the form of $\frac{1}{n}\sum_{t=1}^n g_t(\hat{\theta})$, it is convenient to consider the alternative form $g(\theta_\circ) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n E(g_t(\theta_\circ))$ which is asymptotically equivalent to $g(\hat{\theta})$, where $\theta_\circ$ is the true parameter vector. Following this line of argument, we now consider

$$\nabla d(\theta_\circ) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial d_t(\theta_\circ)}{\partial\theta}\right].$$

Using the normality assumption of the u_t and taking expectation conditional on the information set Ψ_{t-1} iteratively, after some algebra we can get the following simple form of $\nabla d(\theta_\circ)$

$$\nabla d(\theta_\circ) = \begin{bmatrix} 0 & 0 & \nabla d_{13} \\ 0 & 0 & \nabla d_{23} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $\nabla d_{13} = (mx_{11}, \cdots, mx_{kk}, mx_{12}, \cdots, mx_{(k-1)k})'$ is a $\frac{k(k+1)}{2} \times 1$ vector with $mx_{ij} = -\frac{1}{\sigma_u^4}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n(x_{ti} - \underline{x}_{ti}'\phi_\circ)(x_{tj} - \underline{x}_{tj}'\phi_\circ), i,j = 1,2,\cdots,k : i \le j$, and $\nabla d_{23} = (m\varepsilon_{11}, \cdots, m\varepsilon_{pp}, m\varepsilon_{12}, \cdots, m\varepsilon_{(p-1)p})'$ is a $\frac{p(p+1)}{2} \times 1$ vector with $m\varepsilon_{ij} = -\frac{1}{\sigma_u^4}\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \varepsilon_{t-i}\varepsilon_{t-j}, i,j = 1,2,\cdots,k : i \le j$. This implies that $\nabla d(\theta_\circ)$ can be estimated consistently by the matrix $\nabla d(\hat{\theta})$ which is given below

$$\nabla d(\hat{\theta}) = \begin{bmatrix} 0 & 0 & \nabla\hat{d}_{13} \\ 0 & 0 & \nabla\hat{d}_{23} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{B.3}$$

where $\nabla\hat{d}_{13} = (\widehat{mx_{11}}, \cdots, \widehat{mx_{kk}}, \widehat{mx_{12}}, \cdots, \widehat{mx_{(k-1)k}})'$ is a $\frac{k(k+1)}{2} \times 1$ vector with $\widehat{mx_{ij}} = -\frac{1}{n\hat{\sigma}_u^4}\sum_{t=1}^n(x_{ti} - \underline{x}_{ti}'\hat{\phi})(x_{tj} - \underline{x}_{tj}'\hat{\phi}), i,j = 1,2,\cdots,k : i \le j$, and $\nabla\hat{d}_{23} = (\widehat{m\varepsilon_{11}}, \cdots, \widehat{m\varepsilon_{pp}},$

$\widehat{m\varepsilon_{12}}, \cdots, \widehat{m\varepsilon_{(p-1)p}})'$ is a $\frac{p(p+1)}{2} \times 1$ vector with $\widehat{m\varepsilon_{ij}} = -\frac{1}{n\hat{\sigma}_u^4} \sum_{t=1}^{n} \hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}, i, j = 1, 2, \cdots, p : i \leq j$.

By the same argument, we can simplify $A(\hat{\theta})$ as follows:

$$A(\hat{\theta}) = \begin{bmatrix} -\frac{1}{n\hat{\sigma}_u^2} \sum_{t=1}^{n} (x_t - \underline{x}_t'\hat{\phi})(x_t - \underline{x}_t'\hat{\phi})' & 0 & 0 \\ 0 & -\frac{1}{n\hat{\sigma}_u^2} \sum_{t=1}^{n} \hat{\underline{\varepsilon}}_t \hat{\underline{\varepsilon}}_t' & 0 \\ 0 & 0 & -\frac{1}{2\hat{\sigma}_u^4} \end{bmatrix} \tag{B.4}$$

Finally, $\nabla l_t(\hat{\theta}) = \frac{\partial l_t(\hat{\theta})}{\partial \theta}$ is easily given from the Appendix A by

$$\nabla l_t(\hat{\theta}) = \begin{bmatrix} \frac{1}{\hat{\sigma}_u^2} \hat{u}_t (x_t - \underline{x}_t'\hat{\phi}) \\ \frac{1}{\hat{\sigma}_u^2} \hat{u}_t \hat{\underline{\varepsilon}}_t \\ -\frac{1}{2\hat{\sigma}_u^2} + \frac{1}{2\hat{\sigma}_u^4} \hat{u}_t^2 \end{bmatrix} \tag{B.5}$$

For the following discussion, recall the definitions of $\underline{\chi}_t, \underline{\xi}_t, \underline{s}_t$ and $\underline{r}_t$, provided in the main text. From (B.2)-(B.5), $a_t(\hat{\theta})$, which is described below, can be easily derived as

$$a_t(\hat{\theta}) = d_t(\hat{\theta}) - \nabla d(\hat{\theta}) A(\hat{\theta})^{-1} \nabla l_t(\hat{\theta}) = (\hat{a}_{t1}, \hat{a}_{t2}, \hat{a}_{t3}, \hat{a}_{t4}, \hat{a}_{t5}, \hat{a}_{t6})' \tag{B.6}$$

where $\hat{a}_{t1} = \frac{1}{\hat{\sigma}_u^4}(\hat{u}_t^2 - \hat{\sigma}_u^2)\underline{\chi}_t, \hat{a}_{t2} = \frac{1}{\hat{\sigma}_u^4}(\hat{u}_t^2 - \hat{\sigma}_u^2)\underline{\xi}_t, \hat{a}_{t3} = \hat{d}_{t3}, \hat{a}_{t4} = \hat{d}_{t4}, \hat{a}_{t5} = \hat{d}_{t5}$, and $\hat{a}_{t6} = \hat{d}_{t6}$.

Now we establish the block diagonality of the covariance matrix of the IM test, say $V(\theta_\circ)$. It is assumed that all conditions stated in White (1982) are satisfied. Given (B.2)-(B.6) with the normality assumption of the u_t, then $V(\theta_\circ)$ takes the form of

$$V(\theta_\circ) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E \begin{bmatrix} \frac{2}{\sigma_u^4}\underline{\chi}_t\underline{\chi}_t' & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{\sigma_u^4}\underline{\xi}_t\underline{\xi}_t' & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{2\sigma_u^8} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{\sigma_u^4}\underline{s}_t\underline{s}_t' & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2\sigma_u^6}\underline{r}_t\underline{r}_t' & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2\sigma_u^6}\underline{\varepsilon}_t\underline{\varepsilon}_t' \end{bmatrix}_{\theta = \theta_\circ}$$

and its diagonal elements are consistently estimated by $\hat{V}_i, i = 1, 2, \cdots, 6$, stated in the main text. To prove this result, we consider

$$V(\theta_\circ) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E[a_t(\theta_\circ) a_t'(\theta_\circ)] \tag{B.7}$$

In the first stage, we evaluate $E[a_t(\theta_\circ)a_t'(\theta_\circ)]$ conditional on the information set Ψ_{t-1} by using the normality assumption of the u_t and taking expectation iteratively. In the next stage, we use the facts that at $\theta = \theta_\circ$, $E(\underline{\varepsilon}_t) = 0, E(\underline{s}_t) = 0$ and $E(\underline{\xi}_t) = 0$ for all t. Then we have the result. Furthermore, it is worth noting that $E(\underline{\xi}_t\underline{\xi}_t')$ which is related to $ARCH$ specification test can be simplified as a diagonal matrix and $E(\underline{\varepsilon}_t\underline{\varepsilon}_t')$ becomes $\sigma_u^2 I_p$.

Acknowledgement. We are grateful to the participants of the 1988 North American Econometric Society Summer Meeting, in particular the discussant Bruce Hansen, for helpful suggestions. We also wish to express our appreciation to Rob Engle, Hal White, Alastair Hall and Pravin Trivedi for constructive comments on an earlier draft of the paper. All errors, of course, remain our own. Financial support from the Research Board and the Bureau of Economic and Business Research is gratefully acknowledged.

REFERENCES

BERA, A.K. (1986), "Model Specification Test through Eigenvalues," Paper presented at the 1986 North-American Summer Meeting of the Econometric Society, Duke University, Durham.

BERA, A.K. and JARQUE, C.M. (1982), "Model Specification Tests: A Simultaneous Approach," *Journal of Econometrics* , **20**, 59–82.

BERA, A.K. and LEE, S. (1988), "Interaction between Autocorrelation and Conditional Heteroskedasticity: A Random Coefficient Approach," Department of Economics, University of Illinois at Urbana-Champaign, mimeo.

BERA, A.K. and MCKENZIE, C.R. (1986), "Alternative Forms and Properties of the Score Test," *Journal of Applied Statistics*, **13**, 13–25.

BOWMAN, K.O. and SHENTON, L.R. (1975), "Omnibus Contents for Departure from Normality based on b_1 and b_2," *Biometrika* , **62**, 243–250.

BUMB, B.L. and KELEJIAN, H.H. (1983), "Autocorrelated and Heteroscedastic Disturbances in Linear Regression Analysis: A Monte Carlo Study," *Sankhyā (Series B)*, **45**, 257–270.

CHESHER, A.D. (1983), "The Information Matrix Test: Simplified Calculation via a Score Test," *Economics Letters* , **13**, 45–48.

CHESHER, A.D. (1984), "Testing for Neglected Heterogeneity," *Econometrica* , **52**, 865–872.

COX, D.R. (1983), "Some Remarks on Overdispersion," *Biometrika*, **70**, 269–274.

DIEBOLD, F.X. (1986), "Testing for Serial Correlation in the Presence of ARCH," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323–328.

ENGLE, R.F. (1982), "Autoregressive Coditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* , **50**, 987–1007.

EPPS, T.W. and EPPS, M.L. (1977), "The Robustness of Some Standard Tests for Autocorrelation and Heteroscedasticity When Both Problems are Present," *Econometrica* , **45**, 745–753.

GODFREY, L.G. and WICKENS, M.R. (1982), "Tests of Misspecification Using Locally Equivalent Alternative Models," in *Evaluating the Reliability of Macro-Economic Models*, ed. by G. C. Chow and P. Corsi. New York: John Wiley & Sons, pp. 71–103.

GOURIEROUX, C., MONFORT, A., RENAULT, E. and TROGNON, A. (1987), "Generalised Residuals," *Journal of Econometrics, Annals 1987-1*, **34**, 5–32.

HALL, A. (1987), "The Information Matrix Test for the Linear Model," *Review of Economic Studies* , **LIV**, 257–263.

JARQUE, C.M. and BERA, A.K. (1987), "An Efficient Large- Sample Test for Normality of Observations and Regression Residuals," *International Statistical Review*, **55**, 163–172.

LANCASTER, T. (1984), "The Covariance Matrix of the Information Matrix Test," *Econometrica* , **52**, 1051–1053.

TSAY, R.S. (1987), "Conditional Heteroscedastic Time Series Models," *Journal of the American Statistical Association* , **82**, 590–604.

WHITE, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* , **48**, 817–838.

WHITE, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* , **50**, 1–25.

WHITE, H. (1987), "Specification Testing in Dynamic Models," in *Advances in Econometrics: Fifth World Congress, Volume I*, ed. by T.F. Bewley, Cambridge: Cambridge University Press, pp. 1–58.